International Journal on Islamic Applications in Computer Science And Technology, Vol. 11, Issue 3, September 2023, 01-10

Localization and Extraction of Qur'an Verses Using Computer Vision

Abdulrahman Salama[1], Umar Siddiqui[2], Adel Sabour[3], and Mohamed Ali[4]

University of Washington
[1]asalama@uw.edu, [2]umarqr@uw.edu, [3]sabour@uw.edu, [4]mhali@uw.edu

Abstract
Localizing Quranic verses, by detecting the verse bounding boxes, with respect to Quran page images is crucial for UI applications. These applications rely on the user interacting with the verse to view the translation, share the verse, listen to its audio, etc. Moreover, the automatic detection of the verse bounding boxes enables additional image processing and analysis of the Quran pages at the verse level. For these use cases, we need to map the user's click within the image boundary and know which verse is selected. In this paper, we propose a computer vision approach using a Faster RCNN neural network to analyze Quran page images and automatically localize the boundary of every verse with respect to the page. This information can, later on, be fed into various UI applications that allow the user to interact with Quran verses. We train our model and run several experiments on the following narrations: Hafs, Douri, Shubah, Qalon, and Warsh. Our results show 100% accurate detection of all verse boundaries for these narrations.

Keywords: **Computer Vision, Detectron2, Faster RCNN, Quran, Quranic Verse Extraction, Quranic Verse Localization.**

1. Introduction

The Quran has been accessed and read online for many years. Many applications provide features where you can click on a certain verse (ayah) and be able to interact with it. However, the process of providing verse boundaries for such applications has been mainly manual. Several tools were built to facilitate the manual data entry of the verse bounding boxes. The process has been labor-intensive from one side. On the other hand, the process needs to be repeated for any new printing of the Quran and for every narration of the Quran. Most of the online applications are available only for Hafs narration. Other narrations lack such features and support. This is the result of not having a clickable window associated with each verse due to the significant amount of manual effort required to label verses for each narration. The lack of windows that allow for verses to be clicked on severely limits the possibilities of having other narrations published on websites with the same functionality as the Hafs narration.

In this paper, we present QR-Vision, a Verse, and the beginning of Surah localIzatiOn Neural Network system. QR representing the website QuranResearch.org (QuranResearch, 2020). A model that detects the verse markers across different narrations. We have tested our model on the following narrations: Hafs, Shubah, Douri, Qalon, and Warsh using Quran pages taken from the King Fahd Glorious Quran Printing Complex (King Fahd Glorious Quran Printing Complex, 2022). However, we believe the model is generalizable to any narration. After detecting the verse markers, we apply an iterative logic to find all the boxes belonging to each verse in between the verse markers. We also detect the beginning of chapters (surahs) and we mark them so we can find the correct boundaries of the first verse of each chapter (surah). We note that the Quran page images associated with different narrations have differences among them in terms of verse locations and numbering. Figure 1, as an example, illustrates the difference between the narrations of Hafs and Douri in the first page. Verse 7 in Hafs is split into two verses in Douri while Douri narration does not consider the basmalah as separate.

Figure 1. Differences Between Narrations in Verse Numbering and Localization

In this paper, our contributions can be summarized as follows:

- We present QR-Vision, a Quran Visual Extraction Neural Network system that detects verse markers and the beginnings of chapters with high accuracy.
- We conduct experiments utilizing this neural network model to automatically scan different narrations and generate bounding boxes for all verses. The narrations we considered are Hafs, Shubah, Douri, Qalon, and Warsh. Yet we believe the model is extensible to other narrations.
- We publish our training datasets, trained model, inference results, and code online to make it available to advance research in this area (GitHub, 2022).

2. Research Background

In this section, we overview related work along two main lines. First, we overview some research directions that utilize Machine Learning techniques to leverage digital access and to verify authentication of the digital Quran copies. Second, we highlight the Deep Learning Models that we used in our computer vision approach to detect verse boundaries.

2.1 Machine Learning Techniques to Detect and Verify Quran Verses

There are several research directions (Sabbah, 2013; Kurniawan, 2013; Hassan, 2020; Hakak, 2018) that focus on validating the authenticity of the Quran verses that are published online. These directions use a form of a matching algorithm, e.g., Boyer Moore algorithm, for verification purposes. Once verification is complete, a watermarking technique is used for the purpose of tamper identification. There are many other types of research around detecting Quran words and verses from text such as (Sabbah, 2014) using support vector machines, and

(Rafea, 2021) which uses a tree-linked hash table data structure to facilitate the matching process.

In (Jannah, 2020), convolution techniques are used to detect Harfu Jar from Quran images, while in (Rizal, 2015), convolution techniques and Bray Curtis distance are used to detect and classify the different Tajwid rules in Quran images. In this paper, we leverage convolution neural networks (CNNs) to detect and localize verse boundaries within a Quran page.

2.2 Machine Learning Techniques to Detect and Verify Quran Verses

Convolution techniques have been the backbone of recent computer vision neural networks. Convolutional neural networks (CNNs) have witnessed tremendous improvements over the last few years. One of the major improvements in this area has been the introduction of Fast R-CNN (Girshick R. , 2015) and Faster R-CNN (Sun, 2017). Faster R-CNN achieved state-of-the-art results on the COCO dataset (Lin, 2014) which is the standard dataset used in object detection tasks.

Recently, transformer-based neural networks have been increasingly used in the field of computer vision such as ViT (Matt, 2021) and Swin Transformer (Guo, 2021). The Swin Transformer uses a shifted window with limited attention to be able to capture features at different scales to achieve better results on segmentation and detection tasks.

In this paper, we use Faster R-CNN based models provided by the Detectron2 library (Girshick Y. W.-Y., 2019). Detectron2 provides models with near state-of-the-art performance for many computer-vision tasks. It also provides rich content in the form of available resources, documentation, examples, and supported frameworks.

3. System Architecture

Figure 2 illustrates the overall architecture of the QR-Vision System. First, we gather sample images of Quran pages from different narrations. We label these sample pages by manually identifying the verse markers in these pages. These sample pages will be used to train the model. We build and retrain a neural network model to localize verse markers and to detect the beginnings of the chapters. We then use the trained neural network model, to localize all verse markers and the beginnings of the chapters in the entire Quran narration. Based on the localization information from the neural network, we perform simple logic to isolate and detect the boundaries of each verse. In the next sections, we discuss each stage of the localization process in more detail.

3.1 Sample Data Collection and Labeling Module

To start the process, we need to train our neural network model with a sample data set. All images are collected from the King Fahd Quran Complex (King Fahd Glorious Quran Printing Complex, 2022). With the quality of images, and with a training dataset that contains only 15 pages, we are able to achieve high accuracy in the detection of both verse markers and chapter beginnings. Each verse marker and each chapter beginning that appears in the 15 training images are carefully labeled manually by us. Figure 3 illustrates the tool we used to identify and box each verse marker and each chapter beginning. Then, we exported this training dataset in the COCO format (Lin, 2014). The model performs well even with a tiny dataset, as we will see later in the results section.

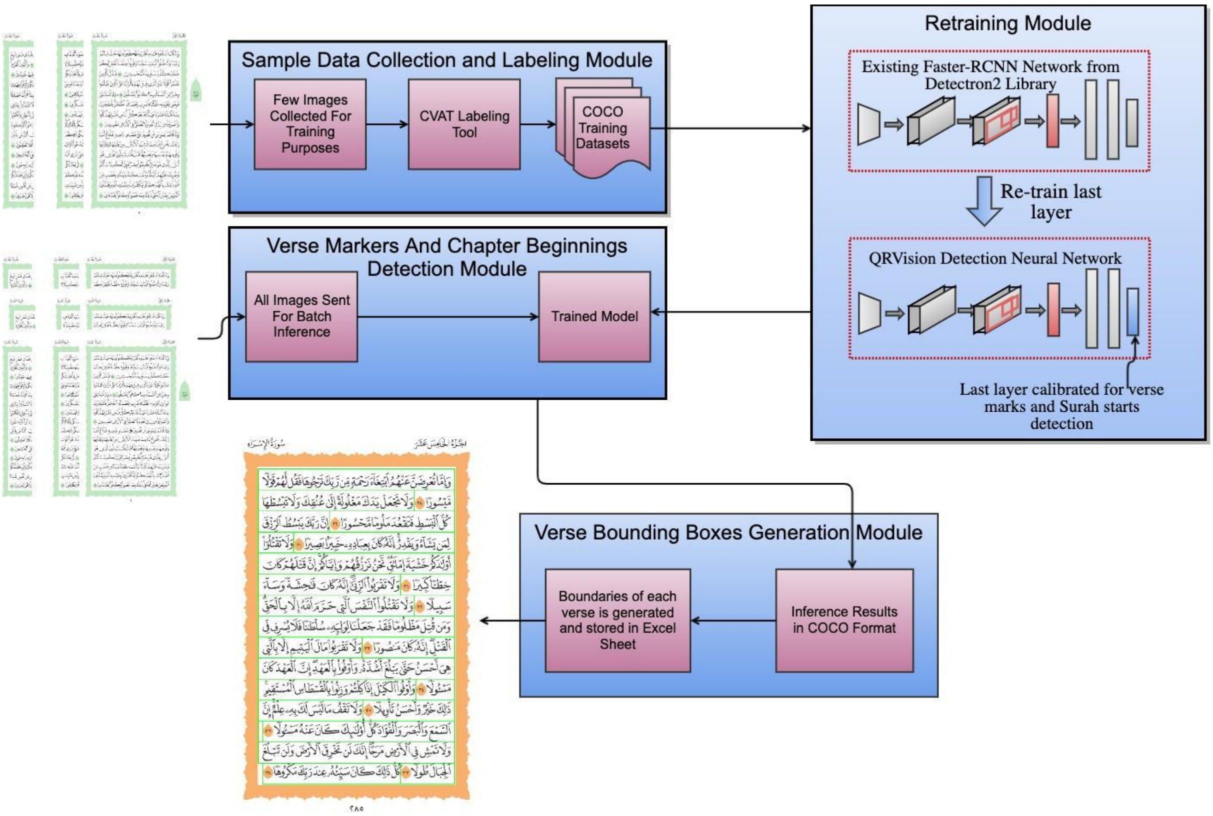


Figure 2. The QR-Vision System Architecture

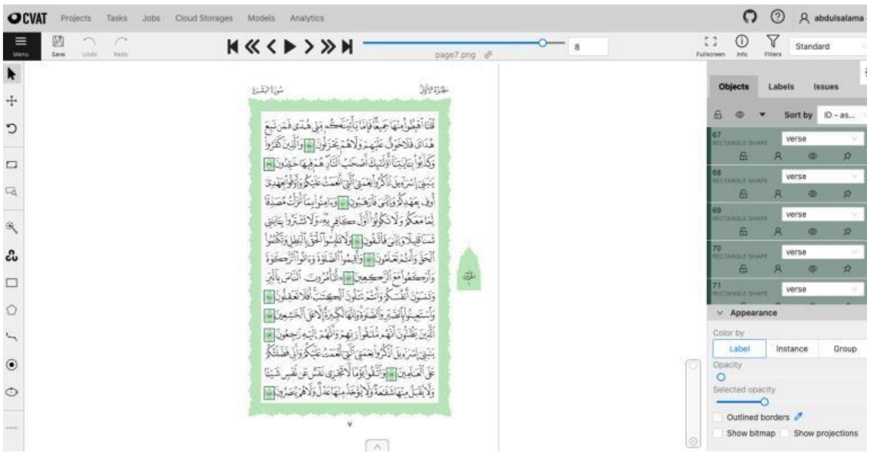


Figure 3a. Labeling Tool for Verse Markers

Figure 3b. Labeling Tool for Chapter Beginnings

3.2 Neural Network Retraining for Visual Extraction

This module uses a convolutional neural network to extract the visual properties of Quran images. Our final goal is to mark verse boundaries and generate boxes representing the area in the image that belongs to each verse separately. In order to achieve this, we first detect and localize verse markers and the beginnings of the chapter. Once we have verse markers, it becomes straightforward to separate the boundaries for each verse. This module focuses on the task of just detecting and localizing verse markers and chapter beginnings. For this purpose, we utilize Detectron2 (Girshick Y. W.-Y., 2019), a library that provides collections of Fast R-CNN (Girshick R., 2015) and Faster R-CNN (Sun, 2017) neural networks for object detection. As illustrated in Figure 4, Faster R-CNN essentially introduces a separate network to learn and predict the region proposals. These regions are fed to the detection network for object detection. The regions make it faster and also more suitable for re-training on different classes since the region proposal network can be re-trained as well.

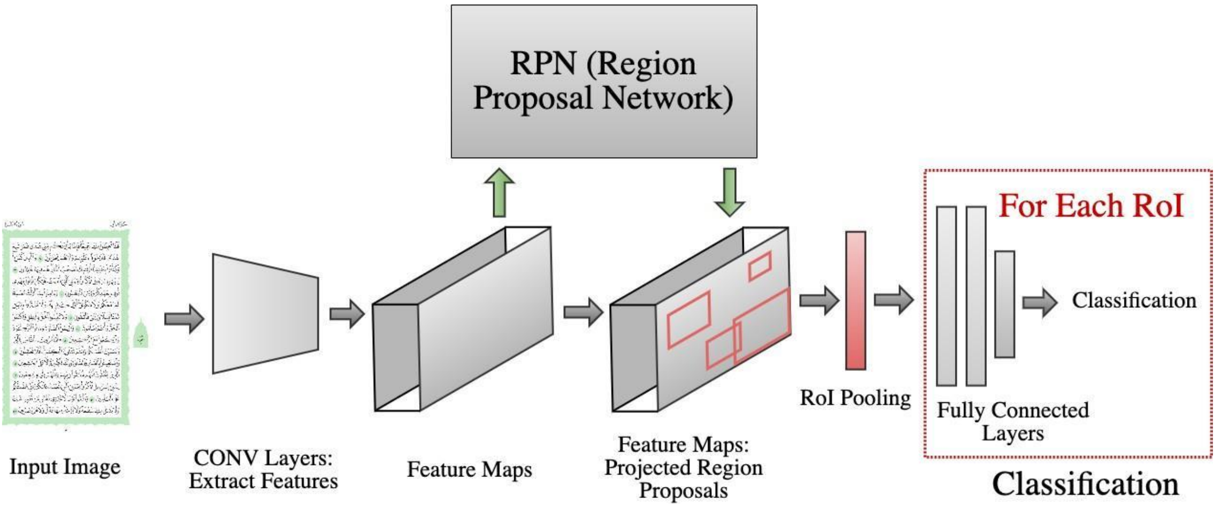


Figure 4. Faster R-CNN High-Level Architecture

We then prepare a new training dataset by labeling a few Quran images obtained by downloading images from the King Fahd Quran Complex. We freeze all layers in the neural network except the last one. The last layer is a fully connected layer. We use the training dataset to retrain the network and tune the weights of this last layer based on the new dataset and desired classes to be detected. We have two classes of objects that we want to detect; verse markers and the beginnings of the chapters (surahs). Our small dataset was enough to retrain the model which shows the power and flexibility of these models. Figure 5 presents the re-training architecture. After retraining and tuning, we use the trained model to do batch inference on all pages of the Quran for verse markers and beginnings of chapter detection.

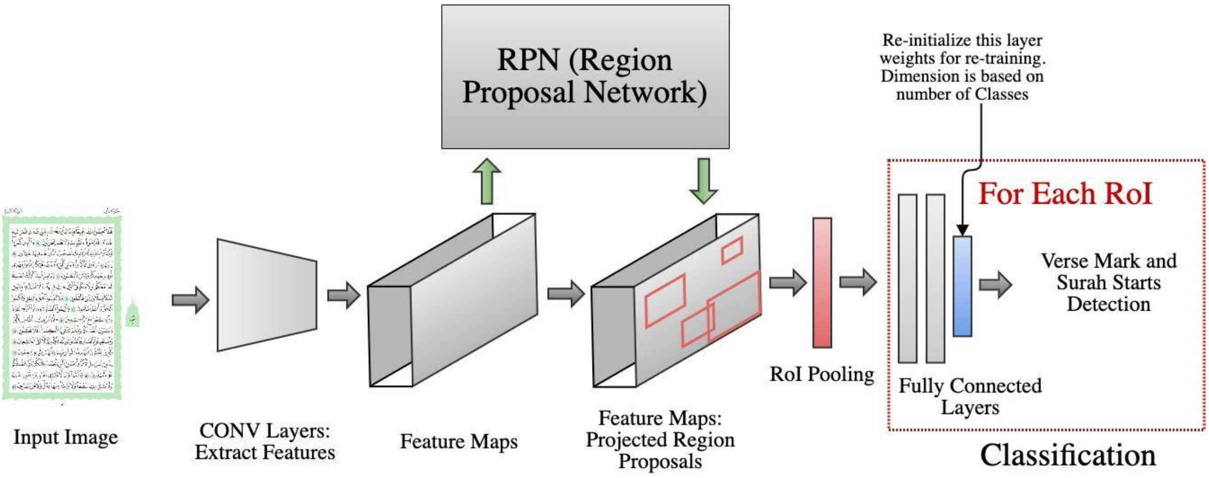


Figure 5. Retraining Architecture at a High Level

3.3 Verse Markers and Chapter Beginnings Detection

This module feeds all Quran images, for a specific narration, into the neural network for batch inference. The neural network processes all pages and detects verse markers and chapter beginnings in all pages and generates boundary boxes around them. The inference results are presented in COCO format. The COCO output file is fed into the next component to finally generate verse boundaries. Figure 6 visualizes the resultant verse markers and chapter beginnings to give a sense of the accuracy of the model.

We analyze the total number of verses and surah beginnings detected by the model and compare them with expected numbers for the narration. After manually checking the output of the model, the number of verses in each Surah was correctly verified.

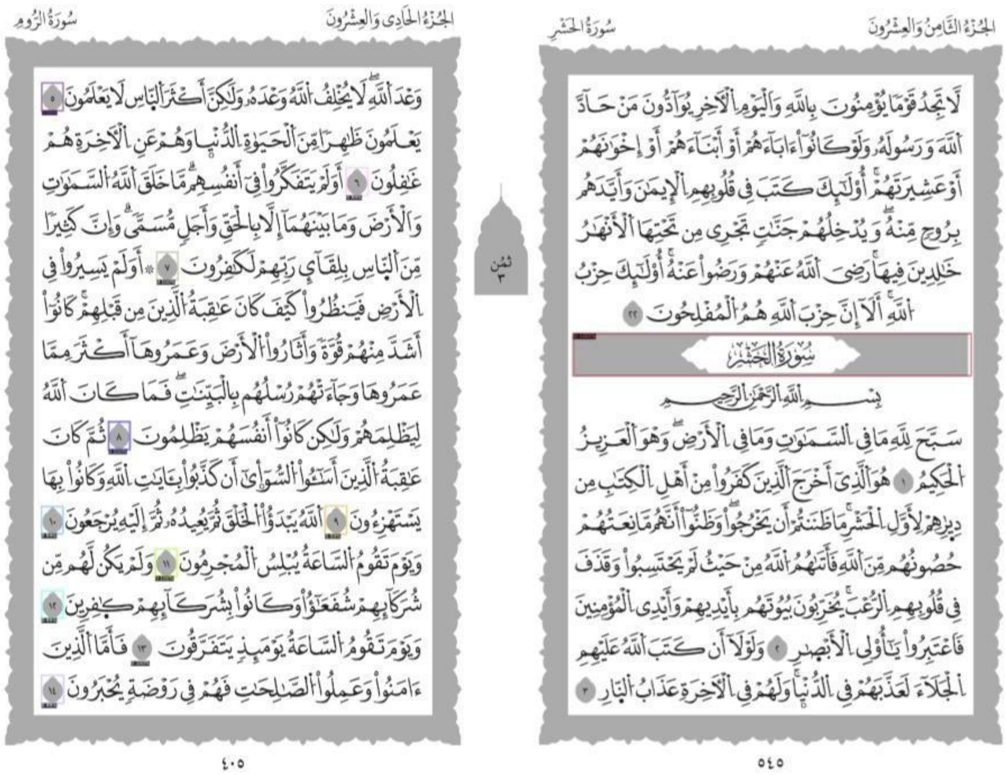


Figure 6. Verse Markers and Chapter Beginnings Detected by The Model

3.4 Bounding Boxes Visualization

In this section, we present the results of generating boxes for Verse boundaries. We use the inference, the verse markers, and chapter beginnings detected by the model, and feed it into the Verse Boundary Generation module. This module processes this information and performs simple logic to draw boxes around the content of each verse. The content belonging to each verse falls in between the current verse marker and the previous verse marker. If the verse is the first verse in the chapter, we use the beginning of the chapter mark to detect the beginning boundary of this verse. Figure 7 illustrates an example.

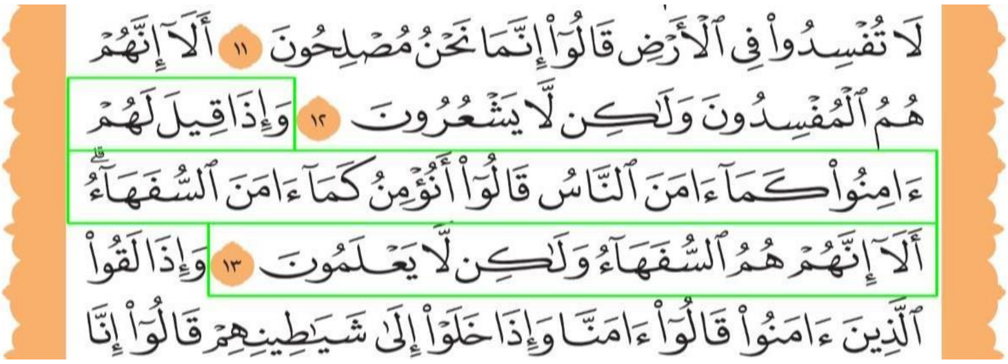


Figure 7. Example Bounding Boxes of a Verse

Figure 7 shows the boundaries of verse 13 of chapter Al-Baqara. The way we achieve this is by using the inference results, which tell us the exact location of verse 13 and the exact location of verse 12. Note that the inference results keep track of the verses in order. With this info, we have a logic that will draw boxes for all the lines in between these two verses. The boundary of each verse can be composed of multiple boxes. In Figure 7, the boundary of verse 13 is presented with a list of 3 boxes. The first bounding box starts from the end of Verse 12 to the end of the line. The second box is the full line in between the two verses. The third box is the box from the beginning of the line until the mark of verse 13.

	A	B	C	D	E	F	G	H
1	Page Number	Verse Serial Number	Box Number	x	y	w	h	
2	1	1	1	808	1099	729	119	
3	1	2	2	590	1099	218	119	
4	1	2	3	712	1218	825	119	
5	1	3	4	590	1218	122	119	
6	1	3	5	1095	1337	442	119	
7	1	4	6	616	1337	479	119	
8	1	5	7	590	1337	26	119	
9	1	5	8	791	1456	746	119	
10	1	6	9	590	1456	201	119	

Figure 8. An Example Tabular Format of the Detected Verse Bounding Boxes

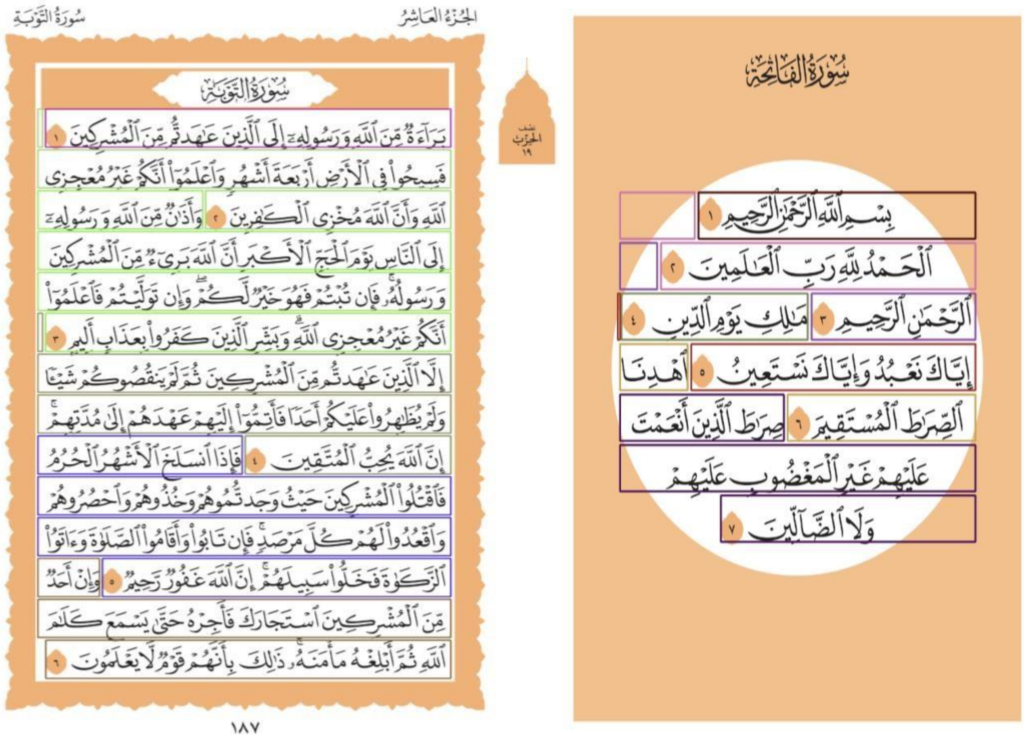
We do this logic iteratively on all verses and mark the boundaries of all verses. We compile and store these results in a table. This format contains for every verse a list of boxes indicating the boundaries that belong to it. Note that we use the verse serial number. Each box is represented with (x,y) coordinates of the top left corner, height, and width. Figure 8 shows a cross-section of this table.

Some of these detected bounding boxes are visualized in Figure 9. For presentational reasons, distinct verse boundaries are indicated with different colors. Figure 10 displays some corner cases that appear in the pages of the Quran. For example, Surah At-Tawba has no Basmallah, and Surah Al-Fatihah has a unique layout across narrations. Despite these exceptions, the model still marks the verses correctly.


سورة البروج الجزء الثلاثون
إلا الذين آمنوا وعملوا الصالحات لهم أجر غير ممنون ٢٥
سورة البروج
بسم الله الرحمن الرحيم
والسماء ذات البروج ١ واليوم الموعود ٢ وشاهد ومشهود ٣
قتل أصحاب الأخدود ٤ النار ذات الوقود ٥ إذ هم عليها
قعود ٦ وهم على ما يفعلون بالمؤمنين شهود ٧ وما نقموا
منهم إلا أن يؤمنوا بالله العزيز الحميد ٨ الذي له ملك
السماوات والأرض والله على كل شيء شهيد ٩ إن الذين
فتنوا المؤمنين والمؤمنات ثم لم يتوبوا فلهم عذاب جهنم ولهم
عذاب الحريق ١٠ إن الذين آمنوا وعملوا الصالحات لهم
جنات تجري من تحتها الأنهار ذلك الفوز الكبير ١١ إن بطش
ربك لشديد ١٢ إنه هو يبدئ ويعيد ١٣ وهو الغفور الودود ١٤
ذو العرش المجيد ١٥ فعال لما يريد ١٦ هل أتاك حديث
الجنود ١٧ فرعون وثمود ١٨ بل الذين كفروا في تكذيب ١٩ والله
من ورائهم محيط ٢٠ بل هو قرآن مجيد ٢١ في لوح محفوظ ٢٢
٥٩٠

سورة الإسراء الجزء الخامس عشر
وإما تعرضن عنهم ابتغاء رحمة من ربك ترجوها فقل لهم قولا
ميسورا ٢٨ ولا تجعل يدك مغلولة إلى عنقك ولا تبسطها
كل البسط فتقعد ملوما محسورا ٢٩ إن ربك يبسط الرزق
لمن يشاء ويقدر إنه كان بعباده خبيرا بصيرا ٣٠ ولا تقتلوا
أولادكم خشية إملاق نحن نرزقهم وإياكم إن قتلهم كان
خطئا كبيرا ٣١ ولا تقربوا الزنى إنه كان فاحشة وساء
سبيلا ٣٢ ولا تقتلوا النفس التي حرم الله إلا بالحق
ومن قتل مظلوما فقد جعلنا لوليه سلطانا فلا يسرف في
القتل إنه كان منصورا ٣٣ ولا تقربوا مال اليتيم إلا بالتي
هي أحسن حتى يبلغ أشده وأوفوا بالعهد إن العهد كان
مسئولا ٣٤ وأوفوا الكيل إذا كلتم وزنوا بالقسطاس المستقيم
ذلك خير وأحسن تأويلا ٣٥ ولا تقف ما ليس لك به علم إن
السمع والبصر والفؤاد كل أولئك كان عنه مسئولا ٣٦
ولا تمش في الأرض مرحا إنك لن تخرق الأرض ولن تبلغ
الجبال طولا ٣٧ كل ذلك كان سيئه عند ربك مكروها ٣٨
٢٨٥

Figure 9. Verse Bounding Boxes Visualization on Page of Quran


سورة التوبة الجزء العاشر
سورة التوبة
براءة من الله ورسوله إلى الذين عاهدتم من المشركين ١
فسيحوا في الأرض أربعة أشهر واعلموا أنكم غير معجزي
الله وأن الله مخزي الكافرين ٢ وأذان من الله ورسوله
إلى الناس يوم الحج الأكبر أن الله بريء من المشركين
ورسوله فإن تبتم فهو خير لكم وإن توليتم فاعلموا
أنكم غير معجزي الله وبشر الذين كفروا بعذاب أليم ٣
إلا الذين عاهدتم من المشركين ثم لم ينقصوكم شيئا
ولم يظاهروا عليكم أحدا فأتموا إليهم عهدهم إلى مدتهم
إن الله يحب المتقين ٤ فإذا انسلخ الأشهر الحرم
فاقتلوا المشركين حيث وجدتموهم وخذوهم واحصروهم
واقعدوا لهم كل مرصد فإن تابوا وأقاموا الصلاة وآتوا
الزكاة فخلوا سبيلهم إن الله غفور رحيم ٥ وإن أحد
من المشركين استجارك فأجره حتى يسمع كلام
الله ثم أبلغه مأمنه ذلك بأنهم قوم لا يعلمون ٦
الحزب ١٩
١٨٧

سورة الفاتحة
بسم الله الرحمن الرحيم ١
الحمد لله رب العالمين ٢
الرحمن الرحيم ٣ مالك يوم الدين ٤
إياك نعبد وإياك نستعين ٥ اهدنا
الصراط المستقيم ٦ صراط الذين أنعمت
عليهم غير المغضوب عليهم
ولا الضالين ٧

Figure 10. Visualization of Some Corner Cases in the Detection of Verse Bounding Boxes

4. Conclusion

In the paper we detected verse markers and the beginnings of the chapters using a neural network model for several different Quran narrations with 100% accuracy. Then, we identified the line spacings in each page to create a list of bounding boxes for each verse. The outcome of this research facilitates the automatic localization of verse boundaries across any Quran narration and any Quran printing. It could also be used to fact-check the number of verses in a Mushaf. Also, several applications can benefit from this research to provide clickable windows around verses for different narrations and Quran printings. We hope that the technologies and processes introduced in this paper can help inspire future work to use robust computer vision techniques to verify new copies of the Quran for authenticity.

5. References

Girshick, R. (2015). Fast r-cnn. Proceedings of the IEEE international conference on computer vision, (pp. 1440-1448). Santiago, Chile: IEEE.

Girshick, Y. W.-Y. (2019). Detectron2. Retrieved May 17, 2022, from https://github.com/facebookresearch/detectron2

Guo, Z. L. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. {IEEE/CVF} International Conference on Computer Vision, {ICCV} 2021 (pp. 9992-10002). Montreal, QC, Canada: IEEE.

Hakak, S. a. (2018). A framework for authentication of digital Quran. (pp. 752--764). N/A: Springer.

Hassan, M. A.-A. (2020). Integrity verification for digital Holy Quran verses using cryptographic hash function and compression. Journal of King Saud University - Computer and Information Sciences, 32(1), 24-34.

Jannah, M. a. (2020). Harfu Jar Detection System In Al-Quran Using Pierce Similarity Algorithm as a Basic Learning Media of Arabic Language. 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), (pp. 349-354). Medan, Indonesia.

King Fahd Glorious Quran Printing Complex 2022. (2022, April 10). Retrieved April 13, 2022, from King Fahd Glorious Quran Printing Complex 2022: https://dm.qurancomplex.gov.sa/en/

Kurniawan, F. a. (2013). Authentication and Tamper Detection of Digital Holy Quran Images., (pp. 291-296). Chengdu, China.

Lin, T.-Y. a. (2014). Microsoft COCO: Common Objects in Context. Computer Vision -- ECCV (pp. 740--755). Cham: Springer International Publishing.

Matt, A. D. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 9th International Conference on Learning Representations, {ICLR} (pp. 3-7).
Austria: OpenReview.net.

QR-Vision Repository. GitHub. (2022). Retrieved August 30, 2022, from https://github.com/WebDataScience/QRVision

Quranresearch.org. QuranResearch. (2020). Retrieved August 30, 2022, from https://quranresearch.org/

Rafea, S. R.-B. (2021). QDetect: An Intelligent Tool for Detecting Quranic Verses in any Text. Procedia Computer Science, 189(1877-0509), 374-381.

Rizal, R. a. (2015). Detection system tajwid al quran on image using bray curtis distance.
International Journal of Computing and Technology, 2(8), 293-300.

Sabbah, T. a. (2013). A Framework for Quranic Verses Authenticity Detection in Online Forum. Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences (pp. 6-11). Madinah, Saudi Arabia: IEEE.

Sabbah, T. a. (2014). Support vector machine-based approach for quranic words detection in online textual content. 8th. Malaysian Software Engineering Conference (MySEC), (pp. 325-330). Langkawi, Malaysia.

Sun, S. R. (2017). Faster {R-CNN:} Towards Real-Time Object Detection with Region Proposal Networks. {IEEE} Trans. Pattern Anal. Mach. Intell., 39(6), 1137-1149.

Biodata

	Abdul Salama is a recent PhD graduate from the School of Engineering and Technology, University of Washington, Tacoma (UWT). Abdul received his Ph.D. degree last summer. Abdul's research focuses on neural network techniques and applications on geospatial Data.
	Umar Ahmed Siddiqui is a student research assistant at the School of Engineering and Technology at the University of Washington, Tacoma. Umar has built the developer’s dashboard of Quranresearch.org website and provided training materials for the prospective developers of the system. Umar’s research interests include AI, Spatiotemporal Databases, and Blockchain technology.
	Adel Sabour is pursuing his PhD in Computer Science and Systems at the School of Engineering and Technology, University of Washington, Tacoma (UWT). He earned his MS from the Department of Computer Science and Information at Cairo University's Institute of Statistical Studies and Research, Egypt. Adel specializes in data mining, machine learning, deep learning, NLP, computer vision, database systems, and big data.
	Mohamed Ali is a professor of Computer Science and Systems at the School of Engineering and Technology, University of Washington, Tacoma (UWT). Mohamed received his Ph.D. degree in computer science from Purdue University. Mohamed’s research interests include Data Science, Big Data Systems, Database Management Systems, and Spatiotemporal Databases.